ENGINEERING AND TECHNOLOGY IN INDIA
Volume 1 Issue 2 (October, 2010) Page : 115-117

REVIEW ARTICLE

Accepted : *June, 2010*

A review of methods used for recognition of Indian language speech

P.P. AGNIHOTRI, R.V. SHINDE AND P.B. KHANALE

ABSTRACT

Automatic Recognition of Voice commands by the machine is a need over a long period of time. In India the language of the peoples changes over every 300 kms. Speech is most complex signal and powerful tool for communication. The Indian Languages are syntactically and semantically different from global language like English. This paper represents the existing methodologies used for recognition of Indian language speech.

Key words : Speech recognition, Speech synthesis, Speaker dependency, Indian languages, HMM

Speech recognition is the process by which a computer identifies spoken words. Basically, it means talking to your computer, and having it correctly recognize what you are saying. The following definitions are the basics needed for understanding speech recognition technology (Cook, 2002).

Utterance:

An utterance is the vocalization (speaking) of a word or words that represent a single meaning to the computer. Utterances can be a single word, a few words, a sentence, or even multiple sentences.

Speaker dependence:

Speaker dependent systems are designed around a specific speaker. They generally are more accurate for the correct speaker, but much less accurate for other speakers. They assume the speaker will speak in a consistent voice and tempo. Speaker independent systems are designed for a variety of speakers. Adaptive systems usually start as speaker independent systems and utilize training techniques to adapt to the speaker to increase their recognition accuracy.

Vocabularies:

Vocabularies (or dictionaries) are lists of words or utterances that can be recognized by the SR system.

Accuracy:

The ability of a recognizer can be examined by measuring its accuracy - or how well it recognizes utterances. This includes not only correctly identifying an utterance but also identifying if the spoken utterance is not in its vocabulary..

Training:

Some speech recognizers have the ability to adapt to a speaker. When the system has this ability, it may allow training to take place. An ASR system is trained by having the speaker repeat standard or common phrases and adjusting its comparison algorithms to match that particular speaker. Training a recognizer usually improves its accuracy.

Automatic Speech recognition system is used to establish the interface between computer system and human beings in which computer system accepts the input as a speech and prints output as text. A survey of research contribution of processing of Indian languages is made [2,3].Chandrashekhar proposed an approach to recognize constant vowels (CU) units in Indian languages using Artificial Neural Network [4].Gangashetty *et al*. represents a neural network model for recognition of syllable like units in Indian languages [5]. Prassanna represents a method to find begin and end points of speech based on vowel onset points (VOP) [6]. Continuous speech recognition system for Hindi [7] describes various improvements carried out to the speech input system and resultant increase in performance.

Methodologies used for recognition of speech are as

Agnihotri, P.P. , Shinde, R.V. and Khanale, P.B. (2010). A review of methods used for recognition of Indian language speech, *Engg. Tech. India*, **1** (2) : 115-117

follows:

Acoustic models:

In a statistical framework, an inventory of elementary probabilistic models of basic linguistic units (e.g., phonemes) is used to build word representations. A sequence of acoustic parameters, extracted from a spoken utterance, is seen as a realization of a concatenation of elementary processes described by hidden Markov models (HMMs). An HMM is a composition of two stochastic processes, a *hidden* Markov chain, which accounts for *temporal* variability, and an observable process, which accounts for *spectral* variability. This combination has proven to be powerful enough to cope with the most important sources of speech ambiguity, and flexible enough to allow the realization of recognition systems with dictionaries of tens of thousands of words[8,9,10].

Word and unit models:

Words are usually represented by networks of phonemes. Each path in a word network represents a pronunciation of the word.

The same phoneme can have different acoustic distributions of observations if pronounced in different contexts. *Allophone* models of a phoneme are models of that phoneme in different contexts. The decision as to how many allophones should be considered for a given phoneme may depend on many factors, e.g., the availability of enough training data to infer the model parameters. [11,12,13]

Language models:

The probability of a sequence of words is computed by a language model (LM). Generation of word hypotheses can result in a single sequence of words, in a collection of the *n-best* word sequences, or in a lattice of partially overlapping word hypotheses. This generation is a search process in which a sequence of vectors of acoustic features is compared with word models. In this section, some distinctive characteristics of the computations involved in speech recognition algorithms will be described, first focusing on the case of a single-word utterance, and then considering the extension to continuous speech recognition[14].

In general, the speech signal and its transformations do not exhibit clear indication of word boundaries, so word boundary detection is part of the hypothesization process carried out as a search. In this process, all the word models are compared with a sequence of acoustic features. In the probabilistic framework, comparison'' between an acoustic sequence and a model involves the computation of the probability that the model assigns to the given sequence. This is the key ingredient of the recognition process.

Conclusion:

Current systems that still need to improvement. The best systems do not perform equally well with different speakers and different speaking environments.

REFERENCES

Cook, Stephen (2002). Speech Recognition HOWTO.

Chandrasekar, M. and Ponnavaikko, M. (2002). A Survey of methods used for Speech processing and the issues related to Indian Language Processing. Int. Conference on Spoken language processing. New Delhi, India, October 2002.

Chandrashekher, M. and Ponnavaikko, M. (2007). Spoken Tamil Character Recognition" Electronic Journal <<Acoustic Technical>> 2007,22

Sekhar, C. Chandra, J. Y. Siva Rama Krishna Rao (1999). Recognition of Consonant-Vowel (CV) units of speech in Indian languages, Proc. National seminar on Information Revolution and Indian Languages, Hyderabad, 22.1–22.6, Nov. 12-14, 1999.

Suryakanth, V. and Gangashetty, B. (2001). Yegnanarayana, Neural network models for recognition of Consonant-Vowel (C^nV) utterances, INNS-IEEE International Joint Conference on Neural Networks, Washington DC, July 14-19, 2001.

Nayeemulla Khan, A., Suryakanth, V., Gangashetty and Rajendran, S. (2002). Speech database for Indian languages - A preliminary study. Proc. Int. Conf. Natural Language Processing, IIT Bombay, Mumbai, 295–301, Dec. 2002.

Samudravijaya, K. Computer Recognition of Spoken Hindi. Proc. Int. Conf. Speech, Music and Allied

Acoustic modeling for speech recognition in telephone based dialog system using limited audio resources. Publisher Springer Berlin / ISSN Heidelberg 0302-9743 (Print) 1611-3349 (Online) Volume 5246/2008

Compact acoustic models for embedded speech recognition. Christophe L´evy, Georges Linar`es, and Jean-Franc¸ois Bonastre, *EURASIP J. on Audio, Speech, and Music Processing* Volume 2009, Article ID 806186, 12 pp.

K. Sri Rama Murty, B. Yegnanarayana, and M. Anand Joseph (2009). Characterization of Glottal Activity From Speech Signals IEEE Signal Processing Letters, **16** (6). JUNE 2009

Zöhre Kara Kayikci and Günther Palm (2008). Word Recognition and Learning based on Associative Memories and Hidden Markov Models. *World Academy of Science, Engineering and Technology,* **37**.

Olsen, Wayne, O., Van Tasell, Dianne, J., Speaks and Charles (1997). Phoneme and Word Recognition for Words in Isolation and in Sentences. *E. Ear & Hearing,* **18** (3) :175-188.

Algorithms for grapheme-phoneme translation for English and French: Applications for database searches and speech synthesis. *Computational Linguistics,* **23** (4) 1997

HMM Methods in Speech Recognition. Renato De Mori & Fabio Brugnara McGill University, Montréal, Québéc,Canada.

Address for correspondence :
P.P. AGNIHOTRI
School of Technology,
Swami Ramand Teerth Marathwada University,
Sub Center, LATURE (M.S.) INDIA
Email : agnihotri_pp@rediffmail.com

Authors' affiliations :
R.V. SHINDE
Department of Computer Science,
Dnyopasak College,
PARBHANI (M.S.) INDIA

P.B. KHANALE
Department of Basic Science,
Marathwada Agricultural University
PARBHANI (M.S.) INDIA

****